Journal of Physiotherapy 60 (2014) 144–150

Journal of PHYSIOTHERAPY

journal homepage: www.elsevier.com/locate/jphys

Research

Several submaximal exercise tests are reliable, valid and acceptable in people with chronic pain, fibromyalgia or chronic fatigue: a systematic review

Julia Ratter [a], Lorenz Radlinger [b], Cees Lucas [c]

[a] Hospital Rivierenland Tiel, The Netherlands; [b] Applied Research and Development Physiotherapy, Health Division, Bern University of Applied Sciences, Switzerland; [c] Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Medical Faculty, University of Amsterdam, Academic Medical Centre, Amsterdam, The Netherlands

KEYWORDS

Chronic pain
Fatigue syndrome (chronic)
Fibromyalgia
Exercise test
Psychometrics
Review (publication type)

ABSTRACT

Question: Are submaximal and maximal exercise tests reliable, valid and acceptable in people with chronic pain, fibromyalgia and fatigue disorders? **Design:** Systematic review of studies of the psychometric properties of exercise tests. **Participants:** People older than 18 years with chronic pain, fibromyalgia and chronic fatigue disorders. **Intervention:** Studies of the measurement properties of tests of physical capacity in people with chronic pain, fibromyalgia or chronic fatigue disorders were included. **Outcome measures:** Studies were required to report: reliability coefficients (intraclass correlation coefficient, alpha reliability coefficient, limits of agreements and Bland-Altman plots); validity coefficients (intraclass correlation coefficient, Spearman's correlation, Kendal T coefficient, Pearson's correlation); or dropout rates. **Results:** Fourteen studies were eligible: none had low risk of bias, 10 had unclear risk of bias and four had high risk of bias. The included studies evaluated: Åstrand test; modified Åstrand test; Lean body mass-based Åstrand test; submaximal bicycle ergometer test following another protocol other than Åstrand test; 2-km walk test; 5-minute, 6-minute and 10-minute walk tests; shuttle walk test; and modified symptom-limited Bruce treadmill test. None of the studies assessed maximal exercise tests. Where they had been tested, reliability and validity were generally high. Dropout rates were generally acceptable. The 2-km walk test was not recommended in fibromyalgia. **Conclusion:** Moderate evidence was found for reliability, validity and acceptability of submaximal exercise tests in patients with chronic pain, fibromyalgia or chronic fatigue. There is no evidence about maximal exercise tests in patients with chronic pain, fibromyalgia and chronic fatigue. **[Ratter J, Radlinger L, Lucas C (2014) Several submaximal exercise tests are reliable, valid and acceptable in people with chronic pain, fibromyalgia or chronic fatigue: a systematic review. *Journal of Physiotherapy* 60: 144–150]**

Introduction

Functional disorders are illnesses in which there is no obvious pathology or anatomical change in an organ, and there is a presumed dysfunction of an organ or system. Chronic pain, fibromyalgia and chronic fatigue disorders are often-mentioned diagnoses belonging to functional disorders.[1] Chronic pain is defined as pain that has lasted longer than 3 to 6 months,[2] although some use 12 months as the threshold.[3] A popular alternative definition of chronic pain, involving no arbitrarily fixed durations is 'pain that extends beyond the expected period of healing'.[2] Fibromyalgia is a chronic functional illness that presents with widespread musculoskeletal pain, including above and below the waist, as well as the right and left sides of the body, and the physical finding of 11 of 18 tender points. These simple criteria provide 85% specificity and sensitivity in differentiating patients with fibromyalgia from those with other rheumatic diseases.[4] Chronic fatigue is defined as persistent or relapsing fatigue lasting more than 6 months, with more than four of the following symptoms: impaired memory, sore throat, tender cervical or axillary lymph nodes, muscle pain, multifocal joint pain, new headaches, unrefreshing sleep, and post-exertion malaise.[4]

A challenging diagnostic dilemma with regard to the above diagnoses is overlap of symptoms. Chronic widespread pain, the cardinal symptom of fibromyalgia, is prevalent and co-occurs with numerous symptom-based conditions such as chronic fatigue syndrome, joint pain and psychiatric disorders.[5] Estimates of the number of patients with fibromyalgia who meet the criteria for chronic fatigue disorders range from 30 to 70%.[4] Fibromyalgia syndrome and chronic fatigue syndrome are similar in many ways – both conditions lack an accepted disease model that can explain signs and symptoms in terms of specific pathophysiological abnormalities.[6]

In Europe, 19% of adults experience chronic pain of moderate to severe intensity with serious negative implications for their social and working lives.[7] Fatigue is also a common symptom in the community, affecting from 0.007 to 2.8% in the general adult population and from 0.006 to 3.0% in primary care.[8] Fibromyalgia syndrome affects 2 to 4% of the general population, and over 5% of patients in general medical practice.[9]

Recent studies have confirmed previous evidence of the enormous indirect socioeconomic costs of chronic pain, fibromyalgia and chronic fatigue disorders. The overall financial costs of chronic pain to society are comparable with the costs of cancer or

http://dx.doi.org/10.1016/j.jphys.2014.06.011

cardiovascular diseases.[10,11] Chronic pain is also associated with many secondary stressors such as sleep disruption, unemployment and interpersonal tensions.[12] Chronic fatigue syndrome is characterised by profound disabling fatigue lasting at least 6 months and accompanied by numerous symptoms such as pain, sleep difficulties and cognitive impairment.[13] Chronic pain, fibromyalgia and chronic fatigue also have personal economic, psychological and social consequences for the affected individuals.[12,14,15] One in three people with pain or fatigue disorders is unable or less able to maintain an independent lifestyle[11] and 50 to 66% of people suffering from chronic pain are less able or unable to exercise, enjoy normal sleep, perform household chores, attend social activities, drive a car, walk or have sexual relations.[16]

Although key risk factors have been identified, the incidence of chronic pain, fibromyalgia and fatigue disorders has been increasing, rendering their management a persistent challenge.[14] Fear avoidance models emphasise psychological distress, pain-related anxiety, anxiety sensitivity, fear of illness/injury, fear of re-injury and catastrophising in the development and maintenance of disabling chronic pain.[17] International and national guidelines recommend graded activity and graded exposure in the treatment of chronic disorders.[15,18–21]

The validity of self-reported assessment of pain and physical disability is controversial. The level of pain reported by people with chronic pain is not always related to their reports of their physical disability. Nevertheless, pain, fear of pain and its consequences are subjective experiences and are difficult to assess.[22] Observational measures may be useful to corroborate subjective reports when evaluating each person's capability.[23,24] Ideally, evaluation of physical function in people with chronic pain and chronic fatigue disorders should rely on a combination of clinical assessment of impairments, behavioural observation of physical function, and self-report.[25] Despite this, there is limited evidence about the acceptability, reliability and validity of submaximal and maximal exercise tests measuring physical fitness and capacity in this group of people. To assess aerobic capacity, maximal testing with calorimetry is considered to be the gold standard.[26,27] However, outcomes of this measurement are strongly influenced by motivation, fear and pain.[26] Furthermore, outcomes are invalid when fear and pain expectation rather than aerobic capacity limit performance.[28] In one study, over 90% of the variance in performance among disabled individuals with chronic musculoskeletal pain was predicted by psychosocial factors like self-efficacy, perceived emotional and physical functioning, pain intensity and pain cognition.[29] Several studies of people with chronic pain have identified discrepancies between self-report of physical activity and actual level of physical activity. Poorer achievement on physical performance testing by people with low back pain has been linked to fear of injury during movement, depression, cognitive factors, pain expectations, pain increase during testing, disability status and the presence of a solicitous spouse.[23]

The conventional Åstrand bicycle test and maximal exercise capacity tests tend to be unacceptable in people with a very poor aerobic capacity[30] and the validity is low in those with chronic low back pain.[27] Also, physical assessments used to detect the degree of disability in other disease states have major limitations when applied to people with fibromyalgia and chronic fatigue syndrome.[31]

In the last decade, many submaximal tests have been developed as an alternative to maximal exercise testing.[28] The most commonly used test in people with chronic low back pain is the submaximal Åstrand bicycle test. Its test-retest reliability seems to be good in people with chronic low back pain.[32] However, submaximal testing tends to underestimate or overestimate maximal oxygen consumption (VO_2max) in 15% of healthy subjects.[33] Nevertheless, due to pain, fatigue and fear of worsening their symptoms, people with chronic pain, fibromyalgia and fatigue disorders are often unable to perform the submaximal Åstrand bicycle test.[34,35]

Guidance for clinicians in this area is needed because the variety in attributes of the available instruments makes it difficult to select the best instrument. Therefore, the research question of this systematic review was:

In people with chronic pain, fibromyalgia and fatigue disorders, are maximal and submaximal physical capacity tests reliable, valid and acceptable?

Method

Identification and selection of trials

A sensitive search was performed in PubMed, Embase, PEDro and the Cochrane library in October 2012. The search strategy was developed by a medical librarian specialist. The detailed strategy for PubMed is presented in Appendix 1 (see eAddenda). Eligible studies could use any study design that reported on one or more measurement properties of physical capacity tests in adults with chronic pain, chronic fatigue disorders or fibromyalgia. Data were extracted for reliability coefficients, validity coefficients and dropout rates. Studies published in any language and in any year were eligible for inclusion.

Records retrieved by the search were assessed for eligibility by two reviewers (JR, LR) working independently, initially based on titles and abstracts, with potentially eligible articles being assessed in full-text to confirm eligibility. Discrepancies were reviewed and consensus was achieved by discussion. Reasons for exclusion were given for each reference and are documented in Figure 1. For each included study, the exercise tests assessed were tabulated along with the psychometric tests performed and their results.

Assessment of characteristics of trials

Quality

The COSMIN 4-point rating scale (excellent, good, fair, poor) was used to evaluate elements of the methodological quality

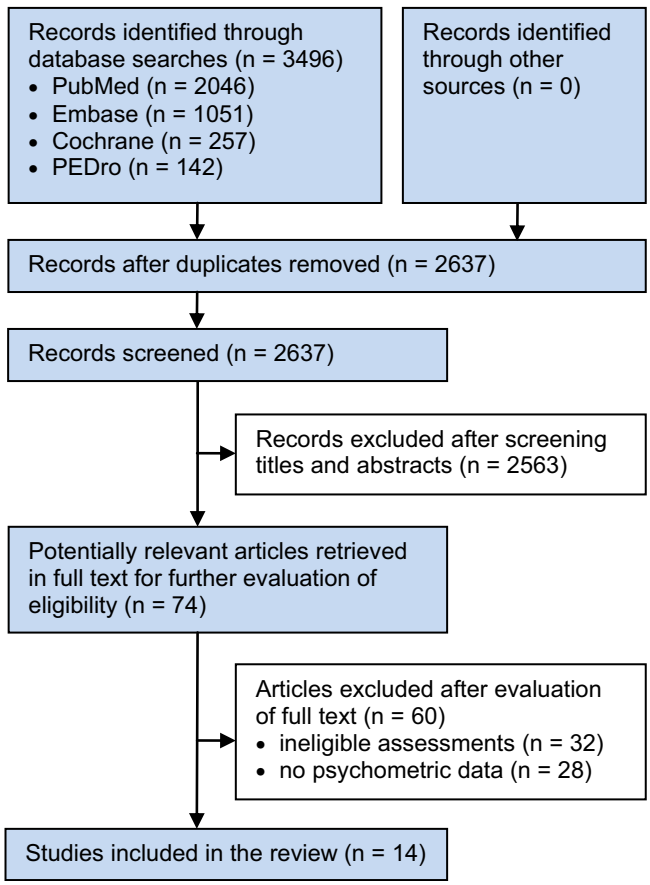

Figure 1. Flow of studies through the review.

(measurement error, reliability, criterion validity) of the studies. A methodological quality score for each relevant element was obtained by taking the lowest rating of any item for that element ('worse score counts').[36] Two authors (JR, LR) independently assessed the risk of bias in included studies, with consensus achieved by discussion.

Participants

Studies involving adults (ie, aged 18 years or older) with chronic pain, fibromyalgia or chronic fatigue disorders were eligible.

Exercise tests

Studies were required to have assessed the psychometric properties of any of the following submaximal exercise tests to be eligible: Åstrand test; modified Åstrand test; Lean body mass-based Åstrand test; submaximal bicycle ergometer test following a protocol other than the Åstrand test; 2-km walk test; shuttle walk test; modified symptom-limited Bruce treadmill test; and walking distance over 5, 6 or 10 minutes.

Psychometric outcomes

Data were extracted, where available, for the following reliability coefficients: intra class correlation (ICC), alpha reliability coefficient, limits of agreements, and Bland-Altman plots. Data were also extracted for the validity coefficients: ICC, Spearman's correlation, Kendal T coefficient, and Pearson's correlation. Dropout rates were also recorded.

Data analysis

The following data were extracted from each eligible study and tabulated: study design, participants (sample size, age, diagnosis), aim, exercise test, psychometric outcomes and methodological quality. Data for individual studies were reported quantitatively and the evidence was also summarised qualitatively. No meta-analyses were performed because of heterogeneity among the study designs used, heterogeneity of the psychometric properties evaluated and incomplete reporting of the data.

The evidence was graded, based on the number of studies, their methodological quality, and the consistency of the available evidence into five categories: strong (consistent findings in two or more high-quality studies); moderate (consistent findings in one high-quality and one low-quality study, or in two or more low-quality studies); limited (only one study); conflicting (inconsistent findings); and no evidence (no studies). The authors considered findings to be consistent if at least 75% of the available studies reported the same conclusion[37].

Results

Flow of trials through the review

The search yielded 3496 records, which amounted to 2637 potentially relevant articles after removal of duplicates. After initial screening, 74 of these articles were obtained in full text for further assessment. The final selection included 14 studies involving 1275 participants. The selection procedure and the reasons for exclusion are presented in Figure 1. Inter-rater agreement about the eligibility of studies was assessed by using an unweighted Kappa. Unweighted Kappa for the selection of abstracts was $k = 0.91$, unweighted Kappa for the selection of full texts was $k = 0.74$; this is considered to be excellent inter-rater agreement.

Characteristics of the included trials

Quality

The reviewers rated the included studies as being 'fair' or 'poor' on the COSMIN checklist, as presented in Appendix 2 (see eAddenda). Common methodological shortcomings were un-blinded assessment, uncertainty about other measurement errors and absence of gold standards.

Participants

Sample sizes in the included studies ranged from 24 to 683. The mean age of all participants was 45 years, with mean age in the individual studies ranging from 34 to 82 years. Age, diagnosis and number of participants in individual studies are presented in Table 1.

Exercise tests

The exercise tests listed above were all assessed by one study each, except for the conventional Åstrand test (three studies), the 5-minute walk test (three studies), and a submaximal bicycle ergometer test following a protocol other than the Åstrand test (three studies).

Are maximal and submaximal physical capacity tests in people with chronic pain, fibromyalgia and fatigue disorders reliable, valid and acceptable?

No data regarding maximal exercise tests in the population of interest were identified. The data extracted from studies of submaximal tests are presented in Table 1. The psychometric properties of each submaximal test are summarised descriptively, below.

Åstrand test

Four studies evaluated the reliability, concurrent validity and dropout rates of the Åstrand test, the modified Åstrand test or the Lean body mass-based Åstrand test. Based on 19 participants, Hodselmans et al reported the test-retest reliability of the Lean body mass-based Åstrand test as an ICC of 0.91 (95% CI 0.76 to 0.97), which changed to 0.96 (95% CI 0.91 to 0.99) when one outlier was excluded.[30] The limits of agreement for the Lean body mass-based Åstrand test were 32.0 and 32.8% including the outlier, and 13.8 and 16.9% excluding the outlier. Assessing the conventional Åstrand test in 31 participants, Keller et al showed a test-retest reliability ICC of 0.96 and a critical difference of 21%.[32] Based on these studies, test-retest reliability seems to be excellent.

Smeets and van Soest evaluated the concurrent validity of the Åstrand test with a modified Åstrand test in 31 participants with musculoskeletal pain disorder.[35] They reported an intraclass coefficient of 0.79 between the two tests. The limits of agreement for VO_2max were 15.9% from the mean difference, which equated to 8.5 ml/kg of lean body mass per minute in VO_2max. Viitanen evaluated the concurrent validity of the Åstrand test with a modified Åstrand test and a 2-km walk test in 69 participants.[39] The ICC was 0.20 (95% CI −0.29 to 0.50) at entry of the study and 0.47 (95% CI 0.15 to 0.67) after 3 months. In addition, Spearman's rank correlation between these two tests was low: $r = 0.37$ ($p < 0.01$) at entry and $r = 0.34$ ($p < 0.01$) after 3 months. These tests showed low and non-significant correlations with the visual analog scale for pain, with r-values ranging from 0.11 to −0.19 for the Åstrand test and 0.09 to −0.22 for the 2-km walk test.

Smeets and van Soest described a slight underestimation of VO_2max with the modified Åstrand test,[35] with VO_2max outcomes an average of 9.96% higher when the conventional Åstrand test was used (95% CI 6.4 to 13.5%) in the pain group. The dropout rates of the Åstrand test and the modified Åstrand test were moderate at 5 to 16%.[30,35] The 2-km walk test was not recommended for subjects with chronic pain syndrome, for example fibromyalgia, due to underestimation of exercise capacity.[38]

Submaximal bicycle ergometer aerobic capacity test not following Åstrand protocol

Three of the 14 studies assessed reliability (test-retest reliability) and acceptability (dropout rate) of other submaximal bicycle ergometer tests. Protocols of these exercise tests are available from the authors. Test-retest reliability was good in the

Table 1
Summary of included studies (n = 14).

Study Design	Participants	Exercise tests	Results	Quality
Smeets[35] Cross-over randomised trial	n = 31 Musculoskeletal pain disorders Age = 48 yr (SD 8)	Åstrand test and modified Åstrand test, repeated at 7 d	Validity: ICC = 0.79, against 33 healthy matched controls. Critical difference: On B-A plots, the limits of agreement = 15.9% (8.5 ml/kg LBM min^{-1}). Slight underestimation of VO_2max with the modified test by 9.96% (95% CI 6.4 to 13.5). Dropout rate: Åstrand test = 5/31 (16%), modified Åstrand test: 5/31 (16%). Usability: modified Åstrand = acceptable	Reliability: fair Criterion validity: fair
Hodselmans[30] Observational study	n = 20 Non-specific chronic LBP with median duration of 68 mth (range 8 to 180) Age = 34 yr (SD 9)	LBM-based Åstrand test, repeated at 2 wk Maximal bicycle test, once for validity	Test-retest reliability: ICC = 0.91 (95% CI 0.76 to 0.97), or excluding outlier 0.96 (95% CI 0.91 to 0.99). Validity: ICC = 0.94 (95% CI 0.85 to 0.98), or excluding outlier 0.98 (955 CI 0.95 to 0.99). Critical difference: On B-A plots, limits of agreement = 32.0 to 32.8%, or excluding outlier 13.8 to 16.9%. Dropout rate: 1/20 (5%)	Reliability: poor
Viitanen[38] Non-randomised intervention study	n = 69 Fibromyalgia Age = 48 yr (SD 7)	Åstrand test and 2-km walk test, repeated at 2 mth	Validity: ICC = 0.20 (95% CI −0.29 to 0.50) at baseline, 0.47 (95% CI 0.15 to 0.67) at 3 mth. Correlation between tests: $r = 0.37$ ($p < 0.01$) at baseline, $r = 0.34$ ($p < 0.01$) at 3 mth. Reliability: alpha coefficients = 0.54 at baseline, 0.47 at 3 mth. Kendal-T coefficients 0.32 and 0.41 for ordinal correlation of the tests. Criterion validity: Results did not correlate with pain VAS, Spearman's $r = -0.19$ to 0.11 for Åstrand test, −0.22 to 0.09 for 2-km walk. Usability: 2-km walk test not recommended for subjects with chronic pain syndrome, eg, fibromyalgia.	Reliability: fair Criterion validity: poor
Keller[32] Observational study	n = 31 LBP Age = 34 yr (range 29 to 43)	Åstrand bicycle test, repeated twice at 5–10 d intervals	Test-retest reliability: ICC = 0.96. Critical difference: On B-A plots, limits of agreement = 21% for LBP.	Reliability: fair
Protas[41] Non-randomised intervention study	n = 683 Chronic work-related spinal disorders Age = 41 yr (SD 9) (cervical); 40 yr (SD 10) (lumbar)	Submaximal bicycle test, assessed before and after rehabilitation	Dropout rate before rehabilitation: 20/179 (11%) cervical, 165/504 (33%) lumbar. Dropout rate after rehabilitation: 0/683 (0%).	Unclear
van Santen[40] Randomised trial	n = 37 Fibromyalgia Age = 42 yr (range 20 to 58)	Bicycle test, repeated at 7 d	Reliability: ICC = 0.86.	Reliability: fair
van Santen[39] Randomised trial	n = 30 Fibromyalgia Age = 45 yr (range 26 to 60)	Bicycle test, repeated at 5 d	Reliability: ICC = 0.70.	Reliability: fair
Smeets[42] Observational study	n = 53 Non-specific chronic LBP without task experience (n = 30) or with task experience (n = 23) Age = 43 yr (SD 9)	5-min walk test, repeated at 5 to 9 d	Test-retest reliability: ICC = 0.89 (95% CI 0.81 to 0.93). Critical difference: On B-A plots, limits of agreement = 20%. Task experience did not significantly influence test-retest differences. Usability: 5-min walk test = acceptable	Reliability: fair

Table 1 (*Continued*)

Study Design	Participants	Exercise tests	Results	Quality
Harding[43] Non-randomised intervention study	n = 147 Chronic pain Age = 50 yr (range 20 to 84)	10-min walk test, measured using 2 raters, and repeated at 12 wk	Inter-rater reliability: ICC = 0.994 (n = 24). Test-retest reliability: Pearson's r = 0.944 (n = 30). Criterion validity: Spearman's r = 0.985, indicating strong correlation with 5-min walk test.	Reliability: poor Criterion validity: fair
Simmonds[45] Observational study	n = 44 LBP Age 43 yr (range 21 to 63)	5-min walk test, repeated at 2 wk	Test-retest reliability: ICC = 0.87. Criterion validity: Pearson's r = 0.617 for correlation between 5-min walk and 50-ft walk. Correlations with 7 other physical performance tests were weak. Usefulness = no specialised equipment required, acceptable to all subjects.	Reliability: fair Criterion validity: poor
Sato[44] Observational study	n = 82 Chronic pain Age = 82 yr (SD 8)	6-min walk test, repeated at 2 wk	Inter-rater reliability: ICC = 1.000. Intra-rater reliability: ICC = 0.979. Validity: Spearman's r = 0.652, indicating significant association with the Functional Independence Measure.	Reliability: fair Criterion validity: poor
Mannerkorpi[46] Observational study	n = 69 Women with fibromyalgia Age = 45 yr (SD 8)	6-min walk test	Dropout rate: 1/69 (1%). 6-min walk test showed a fair relationship with the Physical Function domains of the SF-36 and the FIQ, and a moderate-to-good relationship with the ASES function scale. Correlations between the performance-based tests and the activity limitations tended to be higher than those between performance and pain.	Criterion validity: fair
Taylor[47] Observational study	Reliability: n = 44 Chronic LBP Age = 48 yr (SD 14)	Shuttle walk test, repeated within 1 wk	Reliability: ICC = 0.99. Critical difference: On B-A plots, mean difference = 2.5 m, with limits of agreement of 52 m and −47 m.	Reliability: fair
	Responsiveness: n = 111 Age = 35 (SD 9), 40 (SD 9) and 46 (SD 13) yr	Shuttle walk test, repeated at 2 and 25 mth	Responsiveness: Effect size = 1.42 for fitness group, 0.23 for control group, and 0.94 for orthopaedic clinic group. Usability: simple, quick.	
Wittink[25] Observational study	n = 63 Chronic LBP Age = 40 yr (SD 8)	Modified symptom-limited Bruce-treadmill test	Validity: Spearman's r = 0.43, indicating a strong correlation with the Physical Function domain of the SF-36.	Criterion validity: fair

ASES = American Shoulder and Elbow Surgeons, B-A = Bland-Altman, FIQ = Fibromyalgia Impact Questionnaire, ICC = Intraclass correlation coefficient, LBM = lean body mass, LBP = low back pain, SF-36 = Short-Form Health Survey, VO_2max = maximal aerobic capacity, W_{max} = maximal work capacity.

studies by van Santen et al, [39,40] with ICCs of 0.70 to 0.86. The dropout rates of 0 to 33% among the various tests were considered acceptable.[41]

Walk tests

Five studies evaluated the reliability, criterion validity and acceptability of walk tests. Smeets et al[42] assessed test-retest reliability, reporting an ICC of 0.89 (95% CI 0.81 to 0.93). Harding et al[43] reported a Pearson's r of 0.944. Task experience did not significantly influence test-retest differences.[42] Inter-rater reliability was reported as ICCs of 0.994 by Harding et al[43] and 1.000 by Sato et al.[44] Intra-rater reliability was reported as an ICC of 0.979 by Sato et al[44] and day-to-day reliability as an ICC of 0.87 by Simmonds et al.[45] The critical difference was 20%.[42] Therefore, reliability of the 5-minute, 6-minute or 10-minute walk tests is good to excellent. The 5-minute walk test is considered useful.[42,45] No specialised equipment is required and walk tests appear to be acceptable for people with chronic low back pain.[45]

Criterion validity was established between the 5-minute and 10-minute walk tests with a high Spearman's rank correlation of r = 0.985.[43] Criterion validity of the walk tests was assessed against the 50-foot walk, the Functional Independence Measures (FIM) scale, various performance-based tests, the Short-Form Health Survey (SF-36), the Fibromyalgia Impact Questionnaire (FIQ), and the American Shoulder and Elbow Surgeons (ASES) Function questionnaire. Simmonds et al[45] reported a moderate correlation of the 5-minute walk test with the 50-foot walk, r = 0.617. Sato et al[44] reported a significant correlation of the 6-minute walk test with the Functional Independence Measures scale (r = 0.652, $p < 0.01$), which was used to evaluate activities of daily living. Mannerkorpi et al[46] correlated the 6-minute walk test against various performance-based tests (chair rising test, hand grip strength, endurance shoulder muscles, abduction, hand to neck, hand to scapula) but the criterion validity was fair to moderate, with r-values ranging from −0.46 to 0.63. Criterion validity was established between the 6-minute walk tests and two subscales of the Fibromyalgia Impact Questionnaire: the physical function scale (r = −0.48, $p < 0.001$) and the pain scale (r = −0.39, $p < 0.01$). In the same study,[46] the 6-minute walk test also correlated with the Short-Form Health Survey (SF-36) physical function scale (r = 0.49,

$p < 0.001$), the SF-36 bodily pain scale ($r = 0.38$, $p < 0.01$), the American Shoulder and Elbow Surgeons function scale ($r = 0.68$, $p < 0.001$) and pain scale ($r = 0.53$, $p < 0.001$). In summary, the 6-minute walk test showed a fair relationship with the SF-36 physical function scale and the Fibromyalgia Impact Questionnaire physical function scale, and a moderate-to-good relationship with the American Shoulder and Elbow Surgeons function scale.[46] Concurrent validity with the performance-based tests and the other quality of life scales was low to moderate. The performance-based measures correlated more strongly with activity limitations than with pain.[46] The dropout rate of 1% was low.[46]

Shuttle walk test

Taylor et al[47] reported test-retest reliability with an ICC of 0.99, a mean difference of 2.5 m, and upper and lower limits of agreement of –47 and 52 m. They concluded that the shuttle walk test is a reliable and responsive test and is simple to administer.

Modified symptom-limited Bruce treadmill test

Wittink et al[25] assessed the concurrent validity of the modified treadmill test with the SF-36 scale and found a moderate relationship (Spearman's $r = 0.43$) in 63 people with chronic low back pain.

Discussion

This systematic review identified 14 eligible studies about measurement properties of physical capacity tests in people with chronic pain, fibromyalgia and chronic fatigue disorders. Exhaustive assessment of methodological quality showed some potential bias due to lack of blinding, doubt over whether the measurement was independent, and no gold standard. This may have allowed overestimation of some of the psychometric properties reported. Although the demographic features and disease severity of the participants were comparable among the studies, a meta-analysis could not be performed due to heterogeneity among the study designs used, heterogeneity of the psychometric properties evaluated, and incomplete reporting of the data. Therefore, psychometric data from individual studies were reported quantitatively and qualitatively.

Seven of the 14 studies assessed criterion validity of the submaximal tests with questionnaires or other submaximal tests.[25,35,38,43–46] Difficulties in assessing criterion validity were: low reproducibility, and operationalisation variability of the criterion at issue. However, there is no appropriate reference standard. This could have led to underestimation of the test validity.

None of the included studies mentioned blinding of outcome measurement. This should not have an effect on *reliability* if the test was done in accordance to the test protocol. However, *validity* of the submaximal tests could be overestimated if researchers were aware of the results of the submaximal tests before assessment of the questionnaires. This leads to potential for bias in the review.

The stop criteria of the study protocols were comparable: heart rate too high or too low, signs of serious cardiovascular or pulmonary difficulties, and chest pain. Only one study added 'fatigue' as a stop criterion.[41] This could have led to a higher dropout rate at the entry of the study, compared to results of other studies. It is remarkable, however, that these higher dropout rates are only presented at the start of the study and not at the end. Protas et al[41] hypothesise that this is based on psychosocial fear-avoidance associated with pretesting rather than a true indication of physical deconditioning. Smeets and van Soest[35] suggested strict adherence to the testing protocol and extensive training of the health care providers to increase the acceptability of the exercise tests. Practical experiences show that acceptability of treadmill and bicycle tests is lower in psychosomatic institutions than in outpatient settings. This is attributed to disease severity and other demographic features.

In four of the 14 studies,[38–40,42] assessment of the psychometric properties of the submaximal tests was not the primary purpose of the study. Data of measurement properties were sparse and the methodological shortcomings of the psychometric measurements could have led to bias.

Five out of 14 studies investigated test batteries of physical performance tasks.[42–46] Submaximal exercise tests such as the 5-minute, 6-minute or 10-minute walk tests were merely one item of the test battery. This could have generated an unclear risk of bias and could cause underestimation or overestimation of the effect measure because participants had to do the test battery completely, and not just one exercise test.

Some uncertainties arose about the reliability and criterion validity of the conventional Åstrand test.[27,30,34] Good test-retest reliability (ICC 0.96) was reported in people with chronic low back pain[32] and moderate concurrent validity with the modified Åstrand test (ICC 0.79) in people with musculoskeletal pain disorders.[35] However, the ICC is strongly influenced by the variation between subjects[32] and the low number of participants in the included studies, which may have resulted in a spuriously high estimate of reliability.

Despite good reliability and moderate criterion validity, all the studies showed low levels of perceived exertion. The low levels of perceived exertion may be more likely to be due to fear avoidance than physical deconditioning.

The gold standard for exercise testing is maximal calorimetry, with detailed assessment of lactate, VO_2max, blood pressure and electrocardiographic data. However, these detailed assessments are not available to many physiotherapists. Measuring people's subjective perception with standardised assessment (such as rating of perceived exertion), monitoring heart rate, and performing submaximal exercise tests seem to be the most applicable methods in daily practice. All of the submaximal exercises identified in this review are useful, feasible, and applicable to the population of interest. At most, one session of 20 to 30 minutes is necessary for a submaximal test, although a treadmill or a cycle ergometer are also needed for some of the tests.

Future research in this area should assess the reliability of submaximal exercise tests with higher quality study designs and report data in sufficient detail to allow for meta-analysis. Future studies could also evaluate the concurrent validity of submaximal exercise tests, compared to maximal tests, in people with chronic pain, fibromyalgia and chronic fatigue disorders. However, the lack of studies of maximal testing of people with chronic pain, fibromyalgia and chronic fatigue disorders may be due to difficulties with such tests.[27] Concurrent validity with other physiological measures, such as heart rate variability could also be investigated. Heart rate variability is related to emotional arousal[48] and might be important in the assessment of physical capacity in this population.

In conclusion, there is moderate evidence of the reliability, validity and acceptability of the evaluated submaximal exercise tests in people with chronic pain, fibromyalgia and chronic fatigue disorders. There is no evidence, however, about maximal exercise tests in this population.

What is already known on this topic: Guidelines recommend graded activity in the treatment of chronic pain, fibromyalgia and chronic fatigue disorders. Self-reports of physical disability often do not correlate with pain severity, so objective assessment of physical capacity is recommended.
What this study adds: Although little is known about maximal exercise tests in this population, moderate evidence exists that several submaximal exercise tests are reliable, valid and acceptable in people with chronic pain, fibromyalgia and chronic fatigue disorders.

eAddenda: Appendices 1 and 2 can be found online at doi:10.1016/j.jphys.2014.06.011

Ethics approval: Nil.

Competing interests: There are no conflicts of interests.

Source(s) of support: No sources of support.

Acknowledgements: We are grateful to our friends, family and colleagues.

Correspondence: Julia Ratter, Physiotherapy, Hospital Rivierenland Tiel, The Netherlands. Email: julia.ratter@gmail.com

References

1. Kumar P, Clark M. *Kumar and Clark's Clinical Medicine*. 8th ed. Edinburgh: Saunders; 2012.
2. Turk DC, Okifuji A. Pain terms and taxonomies. In: Loeser D, Butler SH, Chapman JJ, eds. In: *Bonica's Management of Pain.* 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2001:18–25.
3. Main CJ, Spanswick CC. *Pain Management: An Interdisciplinary Approach*. Edinburgh: Churchill Livingstone; 2001:93.
4. Goldenberg DL. Diagnosis and differential diagnosis of fibromyalgia. *Am J Med.* 2009;122:14–21.
5. Kato K, Sullivan PF, Evengård B, Pedersen L. Chronic widespread pain and its comorbidities: A population-based study. *Arch Intern Med.* 2006;166:1649–1654.
6. White KP, Speechley M, Harth M, Ostbye T. Co-existence of chronic fatigue syndrome with fibromyalgia syndrome in the general population: A controlled study. *Scand J Rheumatol.* 2000;29:44–51.
7. O'Connell NE, Wand BM, Marston L, Spencer S, DeSouza LH. Non-invasive brain stimulation techniques for chronic pain-review: A report of a Cochrane systematic review and meta-analysis. *Eur J Phys Rehabil Med.* 2011;47:309–326.
8. Afari N, Buchwald D. Chronic fatigue syndrome: a review. *Am J Psychiatr.* 2003;160:221–236.
9. Berger A, Dukes E, Martin S, Edelsbeerg J, Oster G. Characteristics and healthcare costs of patients with fibromyalgia syndrome. *Int J Clin Prac.* 2007;61:1498–1508.
10. Breivik H, Eisenberg E, O'Brien T. The individual and societal burden of chronic pain in Europe: the case for strategic prioritisation and action to improve knowledge and availability of appropriate care. *BMC Public Health.* 2013;13:1229.
11. International Association for the Study of Pain. EFIC's declaration on chronic pain as a major healthcare problem, a disease in its own right. 2001. http://www.iasp-pain.org [Accessed November 3, 2012].
12. Lillefjell M, Krokstad S, Espnes GA. Prediction of function in daily life following multidisciplinary rehabilitation for individuals with chronic musculoskeletal pain; a prospective study. *BMC Musculoskel Disord.* 2007;8:65.
13. Wyller VB. The chronic fatigue syndrome - an update. *Acta Neurol Scand.* 2007;115: 7–14.
14. Kent P, Mjøsund HL, Petersen DHD. Does targeting manual therapy and/or exercise improve patient outcomes in nonspecific low back pain? A systematic review *BMC Med.* 2010;8:22.
15. Vowles KE, Gross RT, Sorrell JT. Predicting work status following interdisciplinary treatment of chronic pain. *Eur J Pain.* 2004;4:351–358.
16. World Health Organization. World Health Organization supports global effort to relieve chronic pain. 2004. http://www.who.int/mediacentre/news/releases/2004/pr70/en/ [Accessed March 7, 2012].
17. Carleton RN, Abrams MP, Kachur SS, Asmundson GJ. Waddell's symptoms as correlates of vulnerabilities associated with fear-anxiety avoidance models of pain: pain-related anxiety, catastrophic thinking, perceived disability, and treatment outcome. *J Occup Rehabil.* 2009;19:364–374.
18. Edmonds M, McGuire H, Price J. Exercise therapy for chronic fatigue syndrome. *Cochrane Database Syst Rev.* 2004. CD003200.
19. Nijs J, Paul L, Wallman K. Chronic fatigue syndrome: an approach combining self-management with graded exercise to avoid exacerbations. *J Rehabil Med.* 2008;40:241–247.
20. Sumathipala A. What is the evidence for the efficacy of treatments for somatoform disorders? A critical review of previous intervention studies. *Psychosom Med.* 2007;69:889–900.
21. Turk DC. Clinical effectiveness and cost effectiveness of treatments for patients with chronic pain. *Clin J Pain.* 2002;18:355–365.
22. Malliou P, Gioftsidou A, Beneka A, Godolias G. Measurements and evaluations in low back pain patients. *Scand J Med Sci Sports.* 2006;16:219–230.
23. Abdel-Moty AR, Maguire GH, Kaplan SH, Johnson P. Stated versus observed performance levels in patients with chronic low back pain. *Occup Ther Health Care.* 1996;10:3–23.
24. Strong J, Unruh AM, Wright A. *Pain: A Textbook for Therapists.* London: Churchill Livingstone; 2001:123–147.
25. Wittink H, Rogers W, Sukiennik A, Carr DB. Physical functioning: Self-report and performance measures are related but distinct. *Spine.* 2003;28:2407–2413.
26. Nielens H, Cornet D, Rigot F. Maximal versus submaximal testing protocols to evaluate fitness of patients with chronic low back pain. *Med Sci Sports Exerc.* 2002;34(Suppl 1). abstract 270.
27. Wittink H, Hoskins Michel T, Wagner A, Sukiennik A, Rogers W. Deconditioning in patients with chronic low back pain: fact or fiction? *Spine.* 2000;25:2221–2228.
28. Hodselmans AP, Dijkstra PU, van der Schans C, Geertzen JHB. Test-retest reliability of psychophysical lift capacity in patients with non-specific chronic low back pain and healthy subjects. *J Rehabil Med.* 2007;39:133–137.
29. Rudy TE, Lieber SJ, Boston JR, Gourley LM, Baysal E. Psychosocial predictors of physical performance in disabled individuals with chronic pain. *Clin J Pain.* 2003;1: 18–30.
30. Hodselmans AP, Dijkstra PU, Geertzen JHB, van der Schans C. Exercise capacity in non-specific chronic low back pain patients: a lean body mass-based Åstrand bicycle test; reliability, validity and feasibility. *J Occup Rehabil.* 2008;18:282–289.
31. Kelly M, Gagne R, Newman JD, Olney C, Gualtieri C, Trail D. Assessment of fibromyalgia and chronic fatigue syndrome: a new protocol designed to determine work capability: Chronic pain abilities determination (CPAD). *Irish Med J.* 2008;101: 277–278.
32. Keller A, Hellesnes J, Brox JI. Reliability of the isokinetic trunk extensor test, Biering-Sorensen test and Åstrand bicycle test: assessment of intraclass correlation coefficient and critical difference in patients with chronic low back pain and healthy individuals. *Spine.* 2001;26:771–777.
33. Åstrand P, Rohdahl K. *Textbook of work physiology, physiological bases of exercise*. 3rd edn. New York: McGraw-Hill; 1986:354–391.
34. Skinner SJ. Get with the Program: Exercise Prescription for Chronic Fatigue Syndrome and Fibromyalgia. *ACSM's health and fitness journal.* 2005;9:16–21.
35. Smeets RJ, van Soest M. The usability of a modified Åstrand bicycle test to assess the aerobic capacity in patients with musculoskeletal pain and healthy controls. *Disabil Rehabil.* 2009;31:1988–1995.
36. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res.* 2012;21:651–657.
37. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med.* 1997;126:376–380.
38. Viitanen JV. Feasibility of fitness tests in subjects with chronic pain (fibromyalgia): discordance between cycling and 2-km walking tests. *Rheumatol Int.* 2001;21: 1–5.
39. van Santen M, Bolwijn P, Verstappen F, Bakker C, Hidding A, Houben H, et al. A randomized clinical trial comparing fitness and biofeedback training versus basic treatment in patients with fibromyalgia. *J Rheumatol.* 2002;29:575–581.
40. van Santen M, Bolwijn P, Landewé R, Verstappen F, Bakker C, Hidding A, et al. High or low intensity aerobic fitness training in fibromyalgia: does it matter? *J Rheumatol.* 2002;29:582–587.
41. Protas EJ, Mayer TG, Dersh J, Keeley J, Gatchel RJ, McGeary D. Relevance of aerobic capacity measurements in the treatment of chronic work-related spinal disorders. *Spine.* 2004;29:2158–2166.
42. Smeets RJ, Hijdra HJ, Kester AD, Hitters MW, Knottnerus JA. The usability of six physical performance tasks in a rehabilitation population with chronic low back pain. *Clin Rehabil.* 2006;20:989–998.
43. Harding VR, Williams AC, Richardson PH, Nicholas MK, Jackson JL, Richardson IH, et al. The development of a battery of measures for assessing physical functioning of chronic pain patients. *Pain.* 1994;58:367–375.
44. Sato D, Kaneko F, Okamura H. Reliability and validity of the Japanese-language version of the physical performance test (PPT) battery in chronic pain patients. *Disabil Rehabil.* 2006;28:397–405.
45. Simmonds MJ, Olson SL, Jones S, Hussein T, Lee CE, Novy D, et al. Psychometric characteristics and clinical usefulness of physical performance tests in patients with low back pain. *Spine.* 1998;23:2412–2421.
46. Mannerkorpi K, Svantesson U, Broberg C. Relationships between performance-based tests and patients' ratings of activity limitations, self-efficacy, and pain in fibromyalgia. *Arch Phys Med Rehabil.* 2006;87:259–264.
47. Taylor S, Frost H, Taylor A, Barker K. Reliability and responsiveness of the shuttle walking test in patients with chronic low back pain. *Physiother Res Int.* 2001;6: 170–178.
48. Nickel P, Nachreiner F. Sensitivity and diagnosticity of the 0.1-hz component of heart rate variability as an indicator of mental workload. *Hum Factors.* 2003;45: 575–590.